



California Postsecondary Education Commission

Combining Socioeconomic Data with the Commission's Longitudinal Student Data System

Working paper WP/06-06 • September 2006 • www.cpec.ca.gov

The Commission is now using its Longitudinal Student Data System (LSDS) to conduct studies of the movement of students through California's public colleges and universities. The system consists of enrollment and awards data from the University of California, the California State University, and the California Community Colleges. This data is similar to what the segments have been submitting to the Commission since 1976, except that it now includes student identifiers. The identifiers, provided in accordance with AB 1570 (Ch. 916, Statutes of 1999), enable the Commission to conduct longitudinal studies of the progress of individual students and to aggregate student data for evaluation.

Opportunities for linkages to other data sets. A major limitation of the system is that it does not include any socioeconomic data for students, such as parental income or parental education levels. The system has the ethnicity of each student, so the Commission can examine how student progress varies between ethnic groups, but the lack of other student-specific data limits assessing student progress in regard to other characteristics of students.

Although the system does not include socioeconomic information for individual students, the system does include data on the high school the student attended. To some extent, characteristics of a student's high school can be used as a substitute for data on the student's family background. With the existing data, staff cannot compare the progress of students whose parents are university graduates with students who are the first in their family to attend a university, but can, for example, compare the progress of students who attended high school in poorer areas with students who went to high schools in more affluent areas.

Enhancing the high school data. The quality of any analysis using a student's high school depends on the number of records in the data system that have valid high school codes. The segments use a variety of systems to code students' high schools, so not all of the school codes in the data provided by the segments can initially be matched to schools in the Commission's high school database. Other mismatches occur because the codes assigned to individual schools have changed over time or because high school information may not have been collected from students when they enrolled.

For entering freshmen, over three quarters of the records submitted by CSU and UC have school codes that could be immediately matched to definite schools. The Commission maintains an extensive system of translation tables that link different school coding systems and indicate changes in school codes over time. Applying these translations to convert the school codes to a uniform system will increase the number of enrollment records with matchable school codes.

Use of the data. The high school information for entering freshmen is complete enough for the Commission to proceed with analyses showing how the progress of entering freshmen is related to the characteristics of their high school. For example, the Commission could compare retention rates and time-to-degree for students who went to high schools in lower-income areas with progress of students from high schools in more affluent areas. Another possible project is to examine the relationship

between progress in the university systems and the Academic Performance Index (API) of a student's former high school.

Next steps. The investigation already in progress has identified the coding systems that UC and CSU use for foreign and out-of-state schools. Staff has used the results of this investigation to compile translation tables so that these schools will be shown explicitly as out-of-state and foreign schools. At first, many of these schools were unmatchable, because the codes used for foreign and out-of-state schools did not match codes in the Commission's database. In addition, Commission staff has begun a dialog with the California Department of Education to reserve a series of codes for foreign and out-of-state schools to ensure that codes for these schools do not conflict with codes assigned to California high schools.

In the initial data, the high school information is less complete for students entering the four-year systems as juniors. Staff plans to investigate the way that high school data for transfer students is collected and processed. If the high school information for these students can be made more complete, it will be possible to analyze how progress and success of students transferring from community colleges is related to the characteristics of their high school.

A closer look at the 2004 data. To illustrate some of the issues involved in converting school codes to a uniform system, the table opposite shows the 2004 enrollment records for freshmen and juniors. In the initial data, 17% of the freshmen records for UC and 13% of the records for CSU had school codes that did not match schools in the Commission's database. Many of these records are for foreign and out-of-state schools. After conversion to a uniform coding system, these schools will be shown explicitly as foreign and out-of-state schools. The records for juniors have a high percentage of unmatchable codes and show the need to investigate the way information on a student's high school is handled as a student progresses through the system.

The table also shows some anomalies in the way schools are classified. Some of the apparent enrollment by students from elementary schools occurs because schools are classified according to their current status and a former K-12 school may have become an elementary school when a new high school opened in the area.

In the data for the community colleges, 62% of the freshmen records have no information the student's high school. However, many of these records are for noncredit classes where colleges collect less information from students when they enroll. The data for students studying for an associate degree or a certificate is likely to be more complete.

Enrollment records in the initial data

Type of school	Number of records			Percent of records		
	UC	CSU	CCC	UC	CSU	CCC
2004 Freshmen						
Total records	38,257	133,854	1,853,912	100%	100%	100%
Public alternative school	128	472	5,561	—	—	—
Public continuation high school	28	377	20,028	—	—	1
Public middle, junior high	21	14	1,244	—	—	—
Public high or K-12 school	30,696	102,796	571,432	80	77	31
Public adult education	18	86	4,936	—	—	—
Private kindergarten, elementary	169	1,414	985	—	1	—
Private high or K-12 school	839	10,446	36,055	2	8	2
Other identified programs	22	262	1,855	—	—	—
Postsecondary institution	—	1	238	—	—	—
Out-of-state school	—	—	39,877	—	—	2
Foreign school	—	—	20,267	—	—	1
Unmatchable school code	6,336	17,949	—	17	13	—
No school code given	—	37	1,151,434	—	—	62
2004 Juniors						
Total records	61,322	212,278	—	100%	100%	—
Public alternative school	242	172	—	—	—	—
Public continuation high school	203	260	—	—	—	—
Public middle, junior high	2	9	—	—	—	—
Public high or K-12 school	46,360	55,078	—	76	26	—
Public adult education	118	73	—	—	—	—
Private kindergarten, elementary	1,313	1,629	—	2	1	—
Private high or K-12 school	1,547	6,774	—	3	3	—
Other identified programs	14	178	—	—	—	—
Postsecondary institution	5	81	—	—	—	—
Unmatchable school code	11,518	147,896	—	19	70	—
No school code given	—	128	—	—	—	—

Percentages not shown when they round to zero.

